THE ESTATE OF ISABELLA "BELLA" HERNDON, JOHN HERNDON, J.H., *a minor*, T.H., *a minor*, *on behalf of themselves and others similarly situated*, *Plaintiffs*, v. NETFLIX INC., *Defendant*.	Case No. 4:21-cv-6561-YGR **Declaration of Xiao "Bill" Yang in support of Plaintiffs' Motion for Minimal CAFA-Exception-Related Discovery** **Judge:** Hon. Yvonne Gonzalez Rogers

I, Yang Xiao a/k/a Bill Xiao, hereby make this declaration in support of Plaintiffs' Motion for Minimal CAFA-Exception-Related Discovery. I could testify to the following if called by this Honorable Court as a witness and hereby declare that the following statements are true and correct:

1. *Background*: I was contacted by one of Plaintiffs' counsel given my experience and expertise as

 an experienced data scientist to advise regarding the difficulty of providing contact information associated with all accounts that watched a particular show on Netflix.

2. *Experience*: I have over a decade of professional experience working in the field of data analytics and data science, working with large data sets (*i.e.*, Big Data) for large technology companies. In 2015, I joined LinkedIn as a data analyst from a prior position in data analytics at ComScore. After joining LinkedIn, I was quickly promoted to a data-scientist role. From there, I transferred to a new job in data science at Facebook (now Meta) before coming back to LinkedIn. In the course of

over a decade working as a data analyst and then data science, I have had deep and extensive experience working with massive datasets and databases, across many different software programs and platforms, to pull large amounts of data from databases in useable format. In these roles, I developed what are called "queries" which are basically short programs to pull large amounts of data for analysis, often to use that data to advise executives and business decision-makers on my employer's platforms. In my career as a data scientist, I have had many occasions where another employee requested that I pull certain data in a one-off situation While I mostly focused mostly on high level business metrics (like daily active users) and looked at statistics in aggregate, from time to time, I did also look at user engagement at individual level for specific use cases like debugging the tracking data. The type of task of figuring out or identifying a set of users that took some particular action on a large technology company's online platform (here Netflix) is a type of task that would be assigned to a data scientist at these companies.

3.   _Task_: I have reviewed the proposed discovery requests that Plaintiffs' counsel intend to submit alongside the Motion. In essence, I understand that Plaintiffs' counsel want to ask Netflix and/or its employees (likely one of its data scientists) to provide contact information associated with all of the accounts that watched a particular show, _i.e._, "13 Reasons Why."

4.   _Overall Assessment_: Given my experience, the bottom line is that this task would be extremely easy. I recognize that this could be millions of accounts that viewed the show, which may sound like a big number, but that would actually not be a particularly large data pull. Doing it would be easy, not just for an experienced data scientist like me, but I would expect this task to be easy for an entry-level data scientist (or even data analyst with even basic experience pulling data from databases). If I were interviewing a candidate for a data analyst or data scientist role, I would expect questions about how to do this to be easy question Ultimately, I would expect that an entry-level data scientist to complete this task without any real difficulty.

While I have no visibility into how Netflix organize their internal data, I would imagine they store their video viewing log data consisted of time_stamp, user_id, and name_of_video in one table, and user_id, first_name, last_name, email_address, and mailing_address in a separate table. And below is how a basic sample query would look like:

```
SELECT DISTINCT b.first_name, b.last_name, b.email_address, b.mailing_address
```

FROM video_viewing_log_table AS a

JOIN user_information_table AS b

ON a.user_id = b.user_id

WHERE a.name_of_video = '13 Reasons Why';

   The query (which is essentially the short lines of code that would instruct the computer-database systems what data to pull and how to format that data) would take roughly 5-10 minutes to write – less time than it takes to write and review this declaration. If someone were to suggest that this was difficult, I would think that they were either (1) wholly unfamiliar with the basic function of being a data scientist; (2) or lacking in understand of where the necessary data is stored (i.e. someone unfamiliar with which table video logging and user-information data is stored. If data scientists at tech companies are like professors in universities, you wouldn't expect a history professor to answer biology related questions).

;or (3) being intentionally misleading. I cannot stress enough that this task is simple and getting the account contact information associated with accounts that watched a particular show could be readily accomplished by an entry-level data scientist who knew where the data is stored in about 5 to 10 minutes of work, with the rest being time it takes for computer in finish running the query, which might take a while depending on data volume and compute resource. . That entry-level data scientist could accomplish this task writing a query in a few minutes and then sitting back or turning to other tasks while the company's database servers and computers the machines fulfill the instructions

Given that Plaintiffs' counsel want to be asking for this information from a company of the size and technological sophistication of Netflix, I have no doubt that Netflix could easily accomplish this.

5. *Analysis*: Below, I will give further analysis of the task in order to assist the Honorable Court in understanding how a data scientists would approach, but I want to emphasize again that it would take minimal time. This is a truly elementary request. And, I would expect that Netflix has dozens – if not over a hundred personnel – who could accomplish this task with ease.

      a. *Data*:

            i. *Contact   Information*:   Netflix   almost certainly has some database or data repository that houses contact information associated with its user accounts, and it's easiest to think

of it as a big Excel table, where each row is an account and then the columns include a unique account identifier and then remaining columns have contact information. If Netflix didn't have something like this, I would be truly astonished and shocked.

ii. *Viewing History*: Likewise, Netflix almost certainly has some database or data repository that houses viewing history associated with its accounts and, while this database would likely be more complex, it would be very easy to extract which show is being watched by which users. A user can easily find this information from their own Netflix account's viewing history, so we know this information is being stored by Netflix.

iv. *Data Location*: I have not worked for Netflix specifically, but I have worked for large, complex, and sophisticated technology companies like LinkedIn and Facebook / Meta. If I were a full-time employee working in a data-analyst capacity or data-scientist capacity for Netflix and familiar with their data stack, I would expect that I would either know where this data is stored or know who to reach out to figure it out quickly.

b. *Writing the Query*: Then, once I know where the data to be pulled is stored (*i.e.*, "warehoused"), I would write the query to pull it. The query is basically a very short program that instructs the computer systems to put the information from the database and then output it in a specific format, in this case a table or tabular display that would indicate the name and contact information for all accountholders where their account watched "13 Reasons Why."

Again here is the sample query:

SELECT DISTINCT b.first_name, b.last_name, b.email_address, b.mailing_address

FROM video_viewing_log_table AS a

JOIN user_information_table AS b

ON a.user_id = b.user_id

WHERE a.name_of_video = '13 Reasons Why';

Writing this query would take a few minutes for a data analyst or data scientist with basic experience. Essentially, the query would look at the viewing-history data and get the list of accounts by user_id (or any unique identifier associated with the account) that watched a particular show. Then, joining the list of user_ids against a table where user contact information is stored. Again, if I know where the data is stored, writing and executing this query would take me (or even an entry-level data analyst) 5-10 minutes.

c. *Automated Output*: Again, once someone writes the query (which would take a few minutes), the computer systems would do the rest. The only other task would be to figure out a way to provide this data to Plaintiffs' counsel, which would depend upon its file size. But again, whether there were hundreds, thousands, millions, or tens of millions of accounts, the human / manual part would be writing the query and executing it. The remainder would be done by machine – and given what I understand of working for other online platforms like LinkedIn and Facebook, this request would not be a big or onerous request. It seems rather routine and simple.

6. *Compensation*: I was not paid or compensated to provide this testimony. I am not expecting any remuneration of any kind associated with my testimony herein nor have Plaintiffs' counsel offered or promised me any. I am providing this information to help inform the Court.

I declare under penalty of perjury that the foregoing is true and correct.

EXECUTED ON October 22, 2025   Respectfully submitted,

                                            /s/

Yang Xiao a/k/a Bill Xiao

(202) 596-5680

1989xy@gmail.com

Writing this query would take a few minutes for a data analyst or data scientist with basic experience. Essentially, the query would look at the viewing-history data and get the list of accounts by user_id (or any unique identifier associated with the account) that watched a particular show. Then, joining the list of user_ids against a table where user contact information is stored. Again, if I know where the data is stored, writing and executing this query would take me (or even an entry-level data analyst) 5-10 minutes.

c. *Automated Output*: Again, once someone writes the query (which would take a few minutes), the computer systems would do the rest. The only other task would be to figure out a way to provide this data to Plaintiffs' counsel, which would depend upon its file size. But again, whether there were hundreds, thousands, millions, or tens of millions of accounts, the human / manual part would be writing the query and executing it. The remainder would be done by machine – and given what I understand of working for other online platforms like LinkedIn and Facebook, this request would not be a big or onerous request. It seems rather routine and simple.
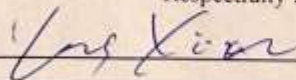
6. *Compensation*: I was not paid or compensated to provide this testimony. I am not expecting any remuneration of any kind associated with my testimony herein nor have Plaintiffs' counsel offered or promised me any. I am providing this information to help inform the Court.

I declare under penalty of perjury that the foregoing is true and correct.

EXECUTED ON October 22, 2025

Respectfully submitted,

/s/ _____

Yang Xiao a/k/a Bill Xiao

(202) 596-5680

1989xy@gmail.com